

IN THE UNITED STATES PATENT OFFICE

Applicant: Kashima, et al.
Serial No.: unassigned
Filed: herewith

Date: May 30, 2001
Docket No.: JP920000069
Group Art Unit: unassigned
Examiner: unassigned

For: Method of Identifying the Source of Genetic Information in DNA

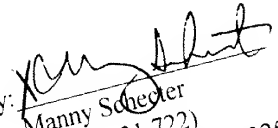
SUBMISSION OF CERTIFIED PRIORITY DOCUMENT

To the Assistant Commissioner of Patents:

Enclosed herewith is a certified copy of the above-identified Application No. 2000-171030, filed in Japan on June 7, 2000, for which applicant claims priority under 35 U.S.C. & 119.

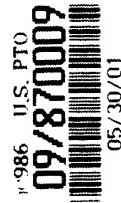
It is requested that this be made part of record in the subject application.

Respectfully submitted,

By: 
Manny Schaefer
(Reg No 31.722)
Telephone: (914) 945-3252

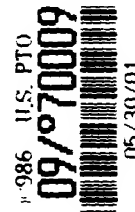
IBM CORPORATION
Intellectual Property Law Dept.
T.J. Watson Research Center
P.O. Box 218
Yorktown Heights, N.Y. 10598

ExpressMail: E-1.761869647US
Date of Deposit: May 30, 2001



JA9-2000 0018

日 本 国 特 許 庁
PATENT OFFICE
JAPANESE GOVERNMENT



別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。
This is to certify that the annexed is a true copy of the following application as filed
with this Office.

出 願 年 月 日
Date of Application:

2000年 6月 7日

出 願 番 号
Application Number:

特願2000-171030

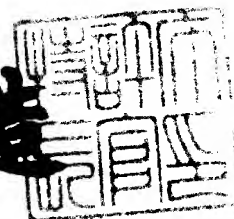
出 願 人
Applicant(s):

インターナショナル・ビジネス・マシーンズ・コーポレーシ
ョン

2000年12月22日

特許庁長官
Commissioner,
Patent Office

及川耕造



出証番号 出証特2000-3107274

【書類名】 特許願

【整理番号】 JP9000069

【提出日】 平成12年 6月 7日

【あて先】 特許庁長官 殿

【国際特許分類】 C12N 15/00
C12N 5/00

【発明者】

【住所又は居所】 神奈川県大和市下鶴間 1 6 2 3 番地 1 4 日本アイ・ビー・エム株式会社 東京基礎研究所内

【氏名】 鹿島 久嗣

【発明者】

【住所又は居所】 神奈川県大和市下鶴間 1 6 2 3 番地 1 4 日本アイ・ビー・エム株式会社 東京基礎研究所内

【氏名】 久保 晴信

【発明者】

【住所又は居所】 神奈川県大和市下鶴間 1 6 2 3 番地 1 4 日本アイ・ビー・エム株式会社 東京基礎研究所内

【氏名】 杉原 亮

【特許出願人】

【識別番号】 390009531

【氏名又は名称】 インターナショナル・ビジネス・マシーンズ・コーポレーション

【代理人】

【識別番号】 100086243

【弁理士】

【氏名又は名称】 坂口 博

【復代理人】

【識別番号】 100104880

【弁理士】

【氏名又は名称】 古部 次郎

【選任した代理人】

【識別番号】 100091568

【弁理士】

【氏名又は名称】 市位 嘉宏

【選任した復代理人】

【識別番号】 100100077

【弁理士】

【氏名又は名称】 大場 充

【手数料の表示】

【予納台帳番号】 081504

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9706050

【包括委任状番号】 9704733

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 DNAへの情報記述方法、遺伝情報の出所認証方法、情報を付加したDNA、塩基配列及び生物体の細胞

【特許請求の範囲】

【請求項1】 DNAの遺伝子以外の部分に通常現れないパターンの塩基配列と、当該DNAが有する所定の遺伝情報の出所を識別するための識別情報とを対応付け、

前記DNAの遺伝子以外の部分に、前記識別情報に対応付けられた前記塩基配列を埋め込むことを特徴とするDNAへの情報記述方法。

【請求項2】 DNAのイントロンに通常現れないパターンの塩基配列と、当該DNAが有する所定の遺伝情報の出所を識別するための識別情報とを対応付け、

前記DNAのイントロンに、前記識別情報に対応付けられた前記塩基配列を埋め込むことを特徴とするDNAへの情報記述方法。

【請求項3】 アミノ酸に翻訳される際のコドンの冗長性を利用して、同一のアミノ酸に翻訳される複数のコドンに2値データに対応付け、

遺伝子のエキソン中に、前記2値データに対応付けられた前記コドンを、所定の情報を示すデータ列となるように配列することを特徴とするDNAへの情報記述方法。

【請求項4】 DNAに遺伝情報の出所を特定する出所識別用の塩基配列を埋め込まれた個体と同種の生物における任意の個体からDNAを取得するステップと、

前記出所識別用の塩基配列に対する相補的な塩基配列を用いて、前記任意の個体の前記DNAに前記出所識別用の塩基配列が存在するかどうかを調べるステップとを含むことを特徴とする遺伝情報の出所認証方法。

【請求項5】 遺伝情報を持つ遺伝子部分と、
遺伝情報を持たない遺伝子以外の部分とを有し、

前記遺伝子以外の部分に、前記遺伝子部分にて伝達される遺伝情報の出所を特定する出所識別情報に対応付けられた塩基配列を持つことを特徴とする情報を付

加したDNA。

【請求項6】 遺伝情報を持つ遺伝子部分に、
蛋白質を合成する際にアミノ酸に翻訳されるエキソンと、
蛋白質を合成する際に切り落とされるイントロンとを有し、
前記イントロンに、前記エキソンに含まれる遺伝情報の出所を特定する出所識別情報に対応付けられた塩基配列を持つことを特徴とする情報を付加したDNA。

【請求項7】 アミノ酸に翻訳される際のコドンの冗長性を利用して2値データに対応付けられた複数種類のコドンを備え、
遺伝子部分における前記コドンの列を、前記2値データにて所定の情報を表すデータ列に対応させたことを特徴とする情報を付加したDNA。

【請求項8】 塩基配列の一部に意図的に設計された特殊配列を含むDNAにおいて、
前記特殊配列は、
DNAが有する遺伝情報の出所を特定する出所識別情報に対応付けられ、
前記DNAが有する遺伝情報の伝達に影響を与えないように前記DNA中に埋め込まれたことを特徴とする情報を付加したDNA。

【請求項9】 前記DNA中の所定の場所に、複数個の前記特殊配列を埋め込んだことを特徴とする請求項8に記載の情報を付加したDNA。

【請求項10】 前記DNA中の所定の場所に、複数種類のパターンを有する前記特殊配列を埋め込んだことを特徴とする請求項8に記載の情報を付加したDNA。

【請求項11】 DNAの一部を構成する塩基配列において、
DNAが有する遺伝情報の出所を特定する出所識別情報に対応付けられ、
前記DNAが有する遺伝情報の伝達に影響を与えないように前記DNA中に埋め込まれたことを特徴とする塩基配列。

【請求項12】 生物体を構成する細胞において、
前記細胞に含まれるDNAは、
遺伝情報を持つ遺伝子部分と、

遺伝情報を持たない遺伝子以外の部分とを有し、

前記遺伝子以外の部分に、前記遺伝子部分にて伝達される遺伝情報の出所を特定する出所識別情報に対応付けられた塩基配列を持つことを特徴とする生物体の細胞。

【請求項 1 3】 生物体を構成する細胞において、

前記細胞に含まれる DNA は、

遺伝情報を持つ遺伝子部分に、

蛋白質を合成する際にアミノ酸に翻訳されるエキソンと、

蛋白質を合成する際に切り落とされるイントロンとを有し、

前記イントロンに、前記エキソンに含まれる遺伝情報の出所を特定する出所識別情報に対応付けられた塩基配列を持つことを特徴とする生物体の細胞。

【請求項 1 4】 生物体を構成する細胞において、

前記細胞に含まれる DNA は、

アミノ酸に翻訳される際のコドンの冗長性を利用して 2 値データに対応付けられた複数種類のコドンを備え、

遺伝子部分における前記コドンの列を、前記 2 値データにて所定の情報を表すデータ列に対応させたことを特徴とする生物体の細胞。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

本発明は、DNA に透かし情報を埋め込むことにより、この DNA に付与された遺伝情報の出所を識別する方法に関する。

【0 0 0 2】

【従来の技術】

種々の動植物には、害虫に強い大豆のような同種の他の個体に比べて優れた特質を持つものが存在する。また、良い血統の競走馬のように人為的な評価基準による価値を持った個体も存在する。これらの特質や価値は、当該個体における遺伝子のレベルで捉えることにより、「付加価値のある遺伝子」として把握することができる。そして今日既に、この付加価値のある遺伝子は、金銭で取り引きさ

れている。例えば、この付加価値のある遺伝子を持った個体は、他の個体に比べて高額で取り引きされる。

【 0 0 0 3 】

ところで、この付加価値のある遺伝子は、自然な交配によって生じる場合もあるが、人工的な操作により意図的に作り出されることが多い。今後、生命科学の進歩に伴い、遺伝子操作により付加価値のある遺伝子（あるいは付加価値のある遺伝子を持った個体）が意図的に生産される事例も増えると考えられる。

この場合、意図的に付加価値のある遺伝子を生産した生産者は、経済的理由などにより、第三者による当該遺伝子の自由な使用を望まない場合がある。例えば、当該遺伝子におけるオリジナルの遺伝情報を持つ生産者が、その使用を一代限りとする、すなわち繁殖や栽培を含む複製、またはDNA（deoxyribonucleic acid：デオキシリボ核酸）やRNA（ribonucleic acid：リボ核酸）のレベルでの複製を禁止するといった条件で、第三者による当該遺伝子の使用を認める場合が考えられる。

【 0 0 0 4 】

しかし、動植物における遺伝情報の複製は、精子や種子を採取することによって、高い技術や高価な装置を必要とせずに行うことが可能である。また、生命工学における技術を用いることにより、DNAやRNAレベルでの遺伝情報の高度な複製を行うことも可能である。

【 0 0 0 5 】

【発明が解決しようとする課題】

上述したように、動植物における遺伝情報の複製は、高い技術や高価な装置を必要とせずに行うことが可能であるし、生命工学における技術を用いれば、DNAやRNAレベルでの遺伝情報の高度な複製を行うことも可能である。

したがって、上述したような付加価値のある遺伝子を第三者が複製することを技術的に制限することは、非常に困難である。

【 0 0 0 6 】

また、所定の生産者により生産された付加価値のある遺伝子が所定の個体から検出された場合に、当該遺伝子が不正に複製されたものかどうかを判断すること

も、突然変異との区別が難しい等の理由で困難である。

【0007】

これに対し、特定の付加価値のある遺伝子におけるDNAの塩基配列の中にID情報などの所定の情報を埋め込むことができれば、当該情報を埋め込まれた遺伝子を複製した場合に当該情報自体も複製されることとなる。このため、当該付加価値のある遺伝子を持つ個体におけるDNAの塩基配列が当該情報を含むかどうかを調べることにより、当該付加価値のある遺伝子が複製されたものかどうかを判断することが可能となる。

【0008】

そこで本発明は、DNAの塩基配列に所定の情報を埋め込むことにより、当該情報を持つDNAにおける遺伝情報の出所を判定できるようにすることを目的とする。

また、本発明は、DNAの塩基配列に意図的に埋め込まれた情報を検出し解析することにより、所定の個体が持つ所定の遺伝子が特定の遺伝子の複製かどうかを判断できるようにすることを他の目的とする。

【0009】

さらにまた、本発明は、所定の個体が持つ所定の遺伝子が特定の遺伝子の複製かどうかを判断可能とすることにより、第三者による特定の遺伝子の不正な複製を防止する手段を提供することを他の目的とする。

【0010】

【課題を解決するための手段】

かかる目的のもと、本発明は、次のように構成されたDNAへの情報記述方法を提供する。すなわち、DNAの遺伝子以外の部分に通常現れないパターンの塩基配列と、このDNAが有する所定の遺伝情報の出所を識別するための識別情報とを対応付け、このDNAの遺伝子以外の部分に、この識別情報に対応付けられた塩基配列を埋め込むことを特徴とする。

ここで、塩基配列のパターンが遺伝子以外の部分に通常現れないとは、このパターンについて、自然な状態でDNAの遺伝子以外の部分に存在しないことが確率的に保証されていることである。この確率は、度数分布を用いた統計的処理に

より算出することができる。

【 0 0 1 1 】

また、本発明は、DNAの遺伝子以外の部分ではなく、遺伝子部分に情報を記述するDNAへの情報記述方法を提供する。すなわち、DNAのイントロンに通常現れないパターンの塩基配列と、このDNAが有する所定の遺伝情報の出所を識別するための識別情報とを対応付け、このDNAのイントロンに、この識別情報に対応付けられた塩基配列を埋め込むことを特徴とする。

ここで、塩基配列のパターンがイントロンに通常現れないとは、このパターンについて、自然な状態でDNAのイントロンの部分に存在しないことが確率的に保証されていることである。この確率は、度数分布を用いた統計的处理により算出することができる。

【 0 0 1 2 】

さらに本発明は、遺伝情報を含むエキソン自体に情報を記述するDNAへの情報記述方法を提供する。すなわち、アミノ酸に翻訳される際のコドンの冗長性を利用して、同一のアミノ酸に翻訳される複数のコドンに2値データに対応付け、遺伝子のエキソン中に、この2値データに対応付けられたコドンを、所定の情報を示すデータ列となるように配列することを特徴とする。

これにより、コドンの列は、遺伝情報と共に2値データにて表現される情報を多重的に含むこととなる。

なお、コドンの冗長性において、どの種類のアミノ酸に翻訳されるかというコドンの作用自体は共通であっても、コドンの使い方は生物の種毎に差があり、通常偏っている。したがって、生物への影響をできるだけ抑えるために、情報を埋め込もうとする生物において使用頻度の高いコドンを選んで2値データに対応付けることが好ましい。

【 0 0 1 3 】

また、本発明は、以上のようにしてDNAに挿入された情報を用いて、所定の生物から取得したDNAにおける遺伝情報の出所を認識する方法を提供する。すなわち、DNAに遺伝情報の出所を特定する出所識別用の塩基配列を埋め込まれた個体と同種の生物における任意の個体からDNAを取得するステップと、この

出所識別用の塩基配列に対する相補的な塩基配列を用いて、この任意の個体の DNA にこの出所識別用の塩基配列が存在するかどうかを調べるステップとを含むことを特徴とする。

【 0 0 1 4 】

さらに、本発明は、上記の情報記述方法を用いて情報を付加した DNA を提供する。すなわち、かかる DNA は、遺伝情報を持つ遺伝子部分と、遺伝情報を持たない遺伝子以外の部分とを有し、この遺伝子以外の部分に、この遺伝子部分にて伝達される遺伝情報の出所を特定する出所識別情報に対応付けられた塩基配列を持つことを特徴とする。

【 0 0 1 5 】

また、遺伝情報を持つ遺伝子部分に、蛋白質を合成する際にアミノ酸に翻訳されるエキソンと、蛋白質を合成する際に切り落とされるイントロンとを有し、このイントロンに、このエキソンに含まれる遺伝情報の出所を特定する出所識別情報に対応付けられた塩基配列を持つことを特徴とする。

【 0 0 1 6 】

さらにまた、アミノ酸に翻訳される際のコドンの冗長性を利用して 2 値データに対応付けられた複数種類のコドンを備え、遺伝子部分におけるコドンの列を、この 2 値データにて所定の情報を表すデータ列に対応させたことを特徴とする。

【 0 0 1 7 】

さらに、塩基配列の一部に意図的に設計された特殊配列を含む DNA において、この特殊配列は、DNA が有する遺伝情報の出所を特定する出所識別情報に対応付けられ、この DNA が有する遺伝情報の伝達に影響を与えないようにこの DNA 中に埋め込まれたことを特徴とする。

【 0 0 1 8 】

そして、これらの情報を付加した DNA は、遺伝子以外の部分やイントロン、エキソンといった対応場所に、複数個の特殊配列（情報に対応付けられた塩基配列）を繰り返し埋め込んだり、複数種類の特殊配列を埋め込んだりすることができる。

複数個の特殊配列を埋め込んだり、複数種類の特殊配列を埋め込んだりするこ

とにより、この特殊配列が交配などの過程を経て自然に破壊されたり、反対に自然に発生したりする確率を下げるができる。

【 0 0 1 9 】

また、本発明は、上述したDNAに情報を付加するために設計される塩基配列、または上述した情報を付加されたDNAを含む生物体の細胞として提供することができる。

【 0 0 2 0 】

【発明の実施の形態】

以下、添付図面に示す実施の形態に基づいてこの発明を詳細に説明する。

まず、本発明の概要について説明する。

本発明は、DNAにID情報などの所定の情報を持った塩基配列を埋め込むことにより、当該塩基配列を持ったDNAの差別化を図る。以下、この所定の情報を持った塩基配列を透かし配列と呼び、透かし配列によって表される情報を透かし情報と呼ぶ。

この透かし配列を、交配や遺伝子操作などによって得られた付加価値のある遺伝子を持つDNAに埋め込んでおくことにより、その後、交配や種々の手法により当該付加価値のある遺伝子を複製した場合に、当該遺伝子における遺伝情報の出所を判定することが可能となる。また、所定の個体のDNAから透かし配列が検出されたならば、当該個体の持つ遺伝子は、予め透かし配列を埋め込んだDNAから複製されたものであり、突然変異などで自然に生じたものではないと判断することができる。このことを利用すれば、付加価値のある遺伝子が複製された場合、その複製が正当な権利を持って行われたものか、不正に行われたものかを判断することができる。

【 0 0 2 1 】

透かし配列を用いる手順を具体的に説明すると、次のようになる。

(1) 優れた遺伝情報Iをもった精子、卵子、あるいは受精卵などの生殖細胞のDNAに対し、透かし配列Wを埋め込む(Wを埋め込まれた精子、卵子をAとする)。

(2) Aは受精し、成長し成体A'になる。

(3) 成体 A' は、自己の遺伝情報の複製を行い、B (精子、卵子など) を作る。

(4) B は受精し、成長し成体 B' になる。

(5) 成体 B' の DNA から透かし配列 W が検出されると、成体 B' は遺伝情報 I の複製を持っていると判断できる。

なお、交配などにより遺伝子の複製を重ねることや、長時間の経過によって遺伝情報 I は劣化する。したがって、DNA から透かし配列 W が検出できなくなるほど劣化したならば、遺伝情報 I に対する価値も劣化しているとみなすことも可能である。

【 0 0 2 2 】

本発明による DNA の差別化は、DNA 中に透かし配列が存在することのみによって実現する。すなわち、当該透かし配列における透かし情報は、当該 DNA を持つ生体に対して何らかの作用を発現することは期待されない。したがって、本発明は、動植物に限らず全ての生物種に対して利用することが可能である。

【 0 0 2 3 】

次に、遺伝子が蛋白質にコードされる過程の概略を述べることにより、細胞内での遺伝情報の形態について説明する。

図 1 は、DNA 中の遺伝子から蛋白質が合成される過程を説明する図である。

DNA の中には、4 種類の塩基、A (アデニン)、T (チミン)、G (グアニン)、C (シトシン) が並んでいる。DNA の 4 つの塩基 (以下、各塩基を頭文字 A、T、G、C で表記する) の配列の中には、蛋白質コード配列とその転写制御情報が貯えられている遺伝子の部分と、遺伝情報が蓄えられていない部分とがある。図 1 に示すように、蛋白質の合成において、転写のプロセスを経ると、該当する蛋白質の遺伝情報の部分のみが、mRNA と呼ばれる中間遺伝物質の状態に写し取られる。高等生物の場合、この mRNA は、最終的にアミノ酸に翻訳されるエキソン部分と途中で切り取られるイントロン部分とからなる (この状態を 1 次 mRNA という)。そして最終的に、イントロン部分が切り落とされ (スプライシング)、最終 mRNA (成熟 mRNA) となる。この最終 mRNA が翻訳のプロセスを経て蛋白質にコードされる。

【 0 0 2 4 】

次に、遺伝情報を複製する方法について説明する。

遺伝情報の複製方法としては、遺伝情報が物理的にどの様に蓄えられているかに応じて、技術的に異なる種々の方法を用いることができる。例えば、DNAの状態のままコピーするならば、繁殖や栽培といった安価かつ容易な方法、遺伝子部分を含むDNAの一部の領域を切り出す方法などがある。また、RNAなどの他の状態から遺伝情報を複製する方法、遺伝子のシーケンスを読み取って遺伝子を合成する方法なども考えられる。

【 0 0 2 5 】

透かし情報を用いて遺伝情報の出所を判定する場合、上述した種々の複製方法による複製に対して透かし情報が耐性を持っていることが必要である。透かし情報が複製に対して耐性を持つとは、遺伝子を複製した場合に、複製された後まで透かし配列が検出可能な形で存在することを意味する。上述したように、遺伝情報の複製方法として技術的に異なる種々の方法が考えられることから、複製に対する耐性を透かし情報に持たせるため、透かし配列のDNAへの埋め込みは、これらの複製方法を想定して行わなければならない。本発明では、後述する透かし配列の安全性も考慮し、次の3種類の方法を提案する。すなわち、

- a) DNAの遺伝子以外の部分に透かし配列を挿入する方法。
- b) 保護しようとする遺伝子のイントロン部分に透かし配列を挿入する方法。
- c) コドンの冗長性を利用して透かし情報を埋め込む方法。

である。

【 0 0 2 6 】

ここで、コドンの冗長性を利用して透かし情報を埋め込む方法について、さらに説明を補足する。

上述したように、DNAはA、C、G、Tの4種類の塩基で構成されるが、DNAからRNAに転写される時に、チミン(T)はウラシル(Uと略す)に置き換わるので、RNAは、A、C、G、Uの4つの塩基が一行に並んだもので構成される。そして、塩基からアミノ酸への翻訳は、A、C、G、Uの列から3文字一組づつのコドン単位として行われる。

【 0 0 2 7 】

図 1 3 は、コドンとそれに対応するアミノ酸（または特別な意味）をまとめた図表である。

図 1 3 に示す図表（以下、コドン表）は、各欄の左にコドン、右に対応するアミノ酸（または特別な意味）を配置してある。アミノ酸は略称を記述しており、それぞれフェニルアラニン（Phe）、ロイシン（Leu）、セリン（Ser）、チロシン（Tyr）、システイン（Cys）、トリプトファン（Trp）、プロリン（Pro）、ヒスチリン（His）、グルタミン（Gln）、アルギニン（Arg）、イソロイシン（Ile）、メチオニン（Met）、トレオニン（Thr）、アスパラギン（Asn）、リシン（Lys）、セリン（Ser）、バリン（Val）、アラニン（Ala）、アスパラギン酸（Asp）、グルタミン酸（Glu）、グリシン（Gly）である。また、「終止」と書いてあるのは、そのコドンがアミノ酸に翻訳される過程を終止させることを意味している。

【 0 0 2 8 】

コドン表を参照すると、コドンとアミノ酸との対応関係は 1 対 1 ではなく、一つのアミノ酸に翻訳されるコドンは複数あることが分かる。この冗長性は、RNA（または転写される前の DNA）のレベルでの配列が異なっても、最終的なアミノ酸のレベルでは同一の物質が合成されることを表している。

生物としてはアミノ酸が正しく合成されれば良いので、DNA や RNA のレベルではこの冗長性の範囲内であれば、コドンを自由に選ぶことができる。このことを利用し、所定の遺伝子を表すコドンを意図的に選択することにより、透かし情報を記述することができる。

【 0 0 2 9 】

次に、透かし配列として用いることができる塩基配列の条件について説明する。所定の塩基配列を透かし配列として DNA に埋め込むためには、当該塩基配列が安全性と証拠性とを備えていることが必要である。

安全性を備えるとは、発現形として意味がないこと、すなわち DNA に透かし配列を挿入することによって生体に影響を与えないことを意味する。安全性の確認手続きは、例えば食品ならば、厚生省の「組み換え DNA 技術応用食品・食品

添加物の安全性評価指針」などの一般的な基準に従って行われる必要がある。

また、この安全性を考慮すると、透かし配列を挿入可能な位置は、DNAの中で生物学的に特別の意味を持たない部分に限られる。したがって、上述したように、DNAに透かし配列を埋め込む場所として、DNAの遺伝子以外の部分、または遺伝子のイントロン部分が選択されることとなる。なお、コドンの冗長性を利用する場合は、コドンの冗長性により安全性が確保されるので、生物学的に意味を持つエキソン部分に透かし情報を埋め込むことが可能である。

【 0 0 3 0 】

証拠性を備えるとは、透かし配列が検出されることが、複製が行われたことを意味するという保証があることである。すなわち、透かし配列と同じ配列が元々DNAの中に存在したり、多少のDNAの変化によって自然に発生したりしてはいけない。この証拠性を実現するには、透かし配列と同じ配列が自然に現れることがないことを確率的に保証する必要がある。このため、透かし配列のサイズ、個数、種類、挿入位置が問題となる。具体的にこれらのパラメータをどのように設定するかは、後述する。

また、コドンの冗長性を利用して透かし情報を埋め込む場合も、透かし情報を記述するために希少なコドンの組み合わせを採用することが必要である。これにより、遺伝子に偶然入った組み合わせではなく、透かし情報として意図的に遺伝子に組み込まれたものであることを示すことができる。

【 0 0 3 1 】

図2は、透かし配列を決定してDNAに埋め込み、検出するまでの全体的な処理の流れを説明するフローチャートである。

図2を参照すると、まずDNAシーケンスデータに基づいて透かし配列を決定する（ステップ201）。透かし配列の決定方法については後で詳述する。次に、対象となる生物のDNAに当該透かし配列を埋め込む（ステップ202）。次に、当該透かし配列を埋め込まれたDNAの安全性の確認を行う（ステップ203）。当該DNAの安全性が確認されたならば、当該DNAを持つ生物が生産され、当該DNAが複製されることとなる（ステップ204）。この後、必要に応じて同種の生物のDNAから透かし配列の検出を行い、当該生物のもつDNA

情報の出所を確認する（ステップ 2 0 5）。

【 0 0 3 2 】

本発明のように付加価値のある遺伝子の不正な複製を防止する技術として、米国特許公報 U S 5 7 2 3 7 6 5 号に開示された技術がある。この技術は、遺伝子操作によって種子の第 2 世代には芽が出ないようにしたものである。この技術が施された作物から自家採種した種子を撒いても発芽しないため、生産者は毎年種を買うこととなるため、種苗開発企業の利益が保護できるようになる。

この技術では、作物が種苗から生長、開花、受粉というように通常の生長過程を経て、種子が成熟し、休眠期間を経ていよいよ種子の中で第 2 世代の生長点や幼葉が作られる際に、遺伝子に組み込まれた毒素遺伝子により毒性を持つ蛋白質が生産され、その種を殺す。

しかし、この技術において、胚芽を殺す毒素遺伝子の拡散や、毒性を持った蛋白質を摂取した際の人体への影響、特にアレルギーの懸念、また、この種子を食べる鳥や昆虫、カビやウイルスなどの微生物へ与える影響は未知である。本発明は、未知の機能を持つ塩基配列を DNA に埋め込むという点では同様ではあるが、少なくとも明らかに毒性がある物質を作り出すことが分かっている塩基配列を埋め込むことはしない。

また、上記の技術では、毒性のある蛋白質を生産するタイミングを制御するために、発芽する際の胚ができる時にスイッチの入るプロモータを用いている。これに対し、本発明は、生物に依存する機能を必要としないため、種々の生物への利用が容易である。

【 0 0 3 3 】

上述したように、DNA に透かし配列を埋め込む方法として、遺伝情報の複製に対する耐性や安全性を考慮して、DNA の遺伝子以外の部分に挿入する方法、遺伝子のイントロン部分に挿入する方法、及びコドンの冗長性を利用して透かし情報を埋め込む方法を提案した。これらの透かし配列は、透かし配列そのものの形態（挿入する場所や配列のサイズなど）に違いがあるので、挿入可能な透かし配列に対する必要条件はそれぞれ異なる。また、複製に対する耐性や実現の容易もそれぞれ異なっている。

以下、各方法における実施の形態について図面を参照して説明する。

【 0 0 3 4 】

〔第 1 の実施の形態〕

まず、DNA の遺伝子以外の部分に透かし配列を挿入する方法における実施の形態を説明する。

本実施の形態においては、DNA 中から透かし配列が検出されさえすれば、当該 DNA の遺伝情報の出所を特定することができる。したがって、透かし配列が挿入される場所はランダムで良く、DNA の遺伝子部分に挿入されてしまっても、透かし配列の機能を損なうことはない。しかしながら、生物の遺伝情報とは無関係な透かし配列が遺伝子部分に挿入されることによって、当該生物に何らかの影響が生じることが考えられる（後述する、他の実施の形態（イントロン部分に透かし配列を挿入する方法、コドンの冗長性を利用して透かし情報を埋め込む方法）を除く）。そのため、DNA の遺伝子以外の部分に透かし配列を挿入することが必要となる。

図 3 は、本実施の形態による透かし配列の挿入方法の概念を説明する図である。

以下、本実施の形態に関して、図 2 に示した（1）透かし配列の決定、（2）DNA への透かし配列の埋め込み、（3）安全性の確認、（4）透かし配列の検出の各ステップと、（5）透かし配列の耐性について説明する。

【 0 0 3 5 】

（1）透かし配列の決定

透かし配列として利用できる塩基配列を決定する。上述したように、この塩基配列は、透かし配列を埋め込む対象である生物の DNA 中に通常現れないようなパターンを持つ配列（以下、単に通常現れない配列などと略す）である。具体的には、次のようにして決定する。

対象である生物の DNA における全塩基数を N とし、その中に塩基数 n の透かし配列を埋め込むことを考える。塩基は 4 種類（A、T、G、C）あるので、塩基数 n の透かし配列 WM の候補は、 4^n 通りの候補の集合 S の中から条件を満たすものを選べば良い。すなわち、

WMの候補 $\in S(n)$ 要素数 $|S(n)| = 4^n$

【0036】

さらに、この中から通常のDNA中に現れる確率の高い配列と特別の意味を持つ配列を $V(n)$ とすると、透かし配列WMは、 $S(n) - V(n)$ の中から選ばば良い。すなわち、

実際に使う透かし配列 $WM \in S(n) - V(n)$

$S(n)$ の要素数は、 n と共に指数関数的に増えるので、短い長さでも十分に通常現れない配列を見つけることができる。

例えば、人間のDNA（以下、ヒトDNA）に長さ n の透かし配列を埋め込む場合を考える。ヒトDNAの長さは約30億塩基なので、ヒトDNA上に長さ n の部分配列は約30億個あることになる。これらの部分配列に偏りはないと仮定すると、長さ n の透かし配列は 4^n 通り表現できる。そこで、約20塩基程度の長さの部分配列を考えれば、30億通りを大きく越える種類の塩基配列を得られる。したがって、粗く見積もっても、長さ30塩基程度の塩基配列を考えれば、十分に通常現れない配列を作ることができる。この中から、制限酵素認識配列やプロモータなどのような生物学的意味を含んでいないものを選んで透かし配列の候補とすることができる。

実際は、DNA中の任意の部分配列には偏りがある。しかし、現在いくつかの生物におけるDNAの配列決定が完了しており、人間を含むその他の生物についても、次々とDNAの塩基配列が解明されていく見通しである。そして、この全塩基配列に基づいて、部分配列の分布を、近似的にはあるが実際に知ることができる。

【0037】

図4は、DNAシーケンスデータを基にしてこの確率の計算を行い、透かし配列の候補を決定する手順を示すフローチャートである。

図 4 を参照すると、まず、透かし配列として選択する塩基配列が対象である生物の DNA において通常現れない配列であることを保証するための確率のしきい値を設定する（ステップ 4 0 1）。次に、DNA シークエンスデータを用いて、所定の塩基配列が DNA 中に現れる確率を計算する（ステップ 4 0 2）。そして、算出された確率がステップ 4 0 1 で設定されたしきい値よりも小さいものを透かし配列の候補とする（ステップ 4 0 3）。

通常現れない配列の確率は、DNA における塩基配列の全体が既知であれば、これに基づいて近似的に計算できる。

【 0 0 3 8 】

所定の塩基配列が DNA 中に通常現れないことを保証するための確率の計算方法を、長さ 6 塩基の透かし配列を決める過程について、簡単な疑似データを用いて説明する。

まず、透かし配列の候補を決めるため、1 つの個体での長さ 6 塩基の塩基配列の度数分布を考える。図 5 は、この塩基配列の度数分布の例を示す図である。この中で、A A A G T C を透かし配列の候補として選ぶ場合は、A A A G T C の頻度が 3 なので、3 よりも多くの数の A A A G T C を DNA に埋め込むこととすれば、この A A A G T C という塩基配列を透かし配列とすることができる。

ただし、透かし配列を埋め込まれた個体と透かし配列を埋め込まれていない個体とが交配すると、1 回の交配によって発生した個体では、減数分裂により透かし配列の数が約半分になってしまう。これに対応するため、透かし配列は複数個埋め込まなければならない。また、突然変異などによって、透かし配列が壊れたり、逆に偶然に透かし配列と同じ塩基配列が発生したりすることも考慮する必要がある。そこで、埋め込むべき透かし配列の数を、突然変異等の理由により塩基配列の出現頻度に個体差があることを考慮して決めることが必要である。

【 0 0 3 9 】

塩基配列の出現頻度の個体差を考慮する方法として、なるべく多くの個体について、DNA のシークエンスデータを集めることにより、塩基配列の出現頻度ごとの個体数を度数分布表とすることができる。図 6 は、上述した塩基配列 A A A G T C の出現頻度に対する個体数の度数分布表の例を示す図である。

図 6 に示す 1 2 個体のサンプルを仮定して説明すると、A A A G T C が 6 個以上含まれる個体は 1 例であり、全体の 8. 3 % となる。したがって、A A A G T C を透かし配列として用い、一つの個体の DNA に当該配列が 6 個以上入るように埋め込んだ場合、図 6 の疑似データの分布から、透かし配列を入れていないはずの個体のうち、8. 3 % の個体から当該透かし配列が検出されることになる。逆に言うと、この場合、塩基配列 A A A G T C は、8. 3 % 程度の誤り確率で透かし配列として機能していると言える。

【 0 0 4 0 】

さらに、1 種類または数種類の配列を、複数の場所に埋め込むことにより、突然変異によって透かし配列と同一の塩基配列が発生する確率、及び透かし配列が壊れてしまう確率を下げるができる。例えば、1 種類の透かし配列をたくさん埋め込んだ場合に、検出された個数分の透かし配列がすべて突然変異によって変化する確率は非常に低くなる。

【 0 0 4 1 】

この様にして求まる確率を使って、目的から要求される確率を満たすことのできる透かし配列を求める。例として、意図的に生産された遺伝子（付加価値のある遺伝子）の出所を透かし配列を用いて特定することにより、当該遺伝子の不正な使用を防止しようとする場合について、保護期間を 1 0 年として疑似データを用いて説明する。

【 0 0 4 2 】

透かし配列を埋め込む個体と同じ種の個体であって、保護期間の 1 0 年間に存在しうる個体の総数を 1 0 0 0 個体と見積もることができたとする。

また、図 6 に示した度数分布より、DNA に塩基配列 A A A G T C を 6 個埋め込んだ場合で、意図的に当該塩基配列を埋め込んでいない個体から 6 個以上の当該配列が検出される確率（誤り確率）が 8. 3 % であった。同様に、他の塩基配列における誤り確率が、塩基配列 A A A G G T を 1 0 個埋め込んだ場合で 0. 0 2 %、塩基配列 A A A G G G を 8 個埋め込んだ場合で 0. 0 0 1 % であるものとする。

【 0 0 4 3 】

透かし配列として、所定の個体のDNAに、塩基配列A A A G T Cを6個、塩基配列A A A G G Tを10個、塩基配列A A A G G Gを8個DNAに埋め込み、独立事象として確率を計算する。この場合、透かし配列を意図的に埋め込んだ個体以外の個体について、これら全ての塩基配列が、与えられた頻度以上で見つかる確率は、

$$8.3 \times 0.02 \times 0.001 = 0.000166 (\%)$$

なので、上述した保護期間の10年間に存在しうる個体の総数である1000個体中には、偶然同じ配列を持っている個体が発生することはほとんど有り得ないと言える。

【0044】

透かし配列を大量に埋め込んだ場合、標的のDNAを制限酵素などによって断片に分け、DNAチップなどを用いて検出することによって、埋め込まれた透かし配列の量を大まかに知ることができる。これによって、埋め込まれた透かし配列の量が統計的な観点から有意に多ければ、透かし配列が挿入されていると推定する方法を取ることができる。

また、複数種類の塩基配列を透かし配列としてDNAに挿入する場合、この透かし配列の組み合わせを管理することにより、DNAに付加する情報量を増やすことが可能となる。

【0045】

(2) DNAへの透かし配列の埋め込み

透かし配列をDNAに埋め込むことは、ベクターを用いて比較的簡単に行うことができる。

ただし、この方法では透かし配列はDNAの適当な場所にランダムに入る。したがって、埋め込み場所を指定することができないため、目的とする遺伝子以外の場所ではなく、遺伝子部分に挿入される場合がある。そこで、後述する安全性の確認を行うことが不可欠である。

【0046】

(3) 安全性の確認

上述したように、ベクターを用いてDNAに透かし配列を埋め込む場合、埋め込む場所を指定することができないため、透かし配列が遺伝子部分に挿入されてしまう可能性がある。したがって、DNAに透かし配列が埋め込まれた個体の安全性を確認する。ここでは、透かし配列がDNAの遺伝子以外の部分に挿入されていることを確認するわけではないが、本実施の形態は、DNA中から透かし配列が検出されさえすれば、その埋め込まれた場所に関わらず、透かし配列としての機能を発揮できるので、個体の安全性が確認できれば十分である。

安全性の基準は、保護しようとする（不正な複製を防止すべき）付加価値のある遺伝子の機能に応じて決められるべきである。これは、社会的合意を必要とすることではあるが、保護対象である付加価値のある遺伝子が社会的に認められるのであるならば、透かし配列も同程度の安全性で認められると考えられる。

【0047】

この安全性の確認には、生物を用いた実験による安全性のテストが行われるべきである。

図7は、透かし配列の安全性を確認する手順を示すフローチャートである。図7を参照すると、まず、透かし配列の候補の中から任意の一つの配列を選択する（ステップ701）。そして、所定の個体のDNAに選択された透かし配列を埋め込み、当該個体の成長を待つ（ステップ702）。成長した個体の安全性をテストし、問題があれば他の透かし配列候補を選択し、同様の作業を行う（ステップ703）。安全性が確認されたならば、当該配列を透かし配列として決定する（ステップ704）。

【0048】

(4) 透かし配列の検出

DNAから透かし配列を検出することは、埋め込んだ透かし配列に相補的な塩基配列を用いて行うことができる。図8は、相補的な塩基配列を用いて透かし配列を検出する様子を示す図である。

図8を参照すると、DNAに透かし配列TTTATTACAが埋め込まれており、この透かし配列に相補的な塩基配列AAATTAATGTを用いて、この透か

し配列を検出している。

また、検索対象であるDNAを取り出し、シーケンサーにかけて当該DNAの塩基配列を読み取ることにより、当該DNA中に透かし配列が埋め込まれていれば、これを検出することができる。図9は、シーケンサーを用いてDNAの塩基配列を解読し、透かし配列A A A T A A T G Tを検出する様子を示す図である。

【0049】

(5) 透かし配列の耐性

透かし情報が複製に対して耐性を持つためには、DNAが複製された場合に、透かし配列も同時に複製され劣化しないことが必要である。DNAにおける遺伝子以外の場所にランダムに挿入された本実施の形態における透かし配列は、繁殖などによるDNA全体の複製、細胞移植による複製、染色体を抽出して行う複製、DNA中の透かし配列を含む領域を特に切り出して行う複製などに対しては、複製されたDNA（またはその一部である塩基配列）中に劣化せずに複製される。すなわち、透かし情報は耐性を持つ。

しかし、透かし配列が含まれる可能性が十分小さくなる程度のサイズでDNAから切り出した部分を複製した場合は、複製された塩基配列中に透かし配列が複製されることはない。したがって、このような複製に対しては、本実施の形態における透かし情報は耐性を持たない。

また、遺伝子から蛋白質が合成される過程で、蛋白質コード領域が転写されてmRNAになっている段階では、遺伝子以外の部分は存在しないので、透かし配列は含まれていないこととなる。そのため、mRNAの状態から遺伝子を複製した場合は、本実施の形態における透かし情報は耐性を持たない。

【0050】

〔第2の実施の形態〕

次に、保護しようとする遺伝子のイントロン部分に透かし配列を挿入する方法における実施の形態を説明する。

上述したように、高等生物の場合、DNAから蛋白質コード領域を転写したmRNAは、最終的にアミノ酸に翻訳されるエキソン部分と途中で切り取られるイ

ントロン部分とからなる（１次 mRNA）。したがって、透かし配列を挿入した個体への影響を考慮すると、蛋白質の合成に用いられないイントロン部分に透かし配列を埋め込むことができる。

【 0 0 5 1 】

本実施の形態によれば、DNA の遺伝子部分に透かし配列を埋め込むので、保護すべき付加価値のある遺伝子そのものに透かし配列を埋め込むことができるという利点がある。

図 1 0 は、本実施の形態による透かし配列の挿入方法の概念を説明する図である。

以下、本実施の形態に関して、図 2 に示した（１）透かし配列の決定、（２）DNA への透かし配列の埋め込み、（３）安全性の確認、（４）透かし配列の検出の各ステップと、（５）透かし配列の耐性について説明する。

【 0 0 5 2 】

（１）透かし配列の決定

透かし配列として利用できる塩基配列を決定する。この塩基配列は、透かし配列を埋め込む対象である生物の DNA の遺伝子部分におけるイントロン部分に通常現れないような配列である。具体的な透かし配列の決定方法は、第 1 の実施の形態において記載した方法と同様である。ただし、第 1 の実施の形態において、DNA の塩基配列全体において通常現れないような配列を用いたのと異なり、本実施の形態では、遺伝子のイントロン部分に通常現れないような配列を用いる。

【 0 0 5 3 】

第 1 の実施の形態において説明したように、透かし配列として使用する塩基配列は、生物学的意味を持たないことが必要である。

また、イントロン部分に通常現れない配列の確率は、対象となる生物において DNA の遺伝子配列が既知であれば、これに基づいて近似的に計算できる。この確率の計算方法は、多数の個体から集めたイントロン部分の配列を対象として、第 1 の実施の形態の場合と同様の度数分布を用いる方法を採用することができる。

さらに、１種類または数種類の透かし配列を複数のイントロン部分に埋め込むことによって、突然変異によって透かし配列と同一の塩基配列が発生する確率、

及び透かし配列が壊れてしまう確率を下げるができる。そして、複数種類の塩基配列を透かし配列として挿入する場合、この透かし配列の組み合わせを管理することにより、DNAに付加する情報量を増やすことが可能となる。

【 0 0 5 4 】

(2) DNA への透かし配列の埋め込み

遺伝子（エキソン及びイントロンを含む）を合成する過程で、イントロン部分における所望の位置に、透かし配列として決定された塩基配列を挿入することにより、当該透かし配列をDNAに埋め込むことができる。ここで、合成される遺伝子は、保護すべき付加価値のある遺伝子とすることが好ましいが、他の遺伝子部分であっても良い。

また、遺伝子におけるイントロンの場所を特定して塩基配列を埋め込むベクターを用いることができれば、所望の遺伝子のイントロン部分に透かし配列を埋め込むことができる。

なお、本実施の形態により透かし配列の埋め込みを行うためには、遺伝子におけるイントロン部分を見つけることが必要である。遺伝子からイントロンを切り取るスプライシングは、スプライソームRNAによって引き起こされているが、その理由は、スプライソームRNAが持つ塩基配列がイントロン開始部分の塩基配列と結合し易いためであることが分かっている。そこで、このスプライソームRNAが持つ塩基配列を利用して遺伝子のイントロン部分を特定することが考えられる。

【 0 0 5 5 】

(3) 安全性の確認

上述したように、イントロンは遺伝子がアミノ酸に翻訳される際にスプライシングにより切り取られる部分である。しかしながら、遺伝子部分に生物の遺伝情報とは無関係な透かし配列が挿入されることから、本実施の形態においても、透かし配列が埋め込まれた個体の安全性を確認することが必要である。

安全性の基準は、上述した第 1 の実施の形態と同様に、保護しようとする付加価値のある遺伝子の機能に応じて決められるべきである。

また、安全性の確認には、生物を用いた実験による安全性のテストが行われる

べきである。透かし配列の安全性を確認する手順は、第 1 の実施の形態における図 7 に示した手順と同様である。

【 0 0 5 6 】

(4) 透かし配列の検出

透かし配列の検出は、第 1 の実施の形態の場合と同様に、埋め込んだ透かし配列に相補的な塩基配列を用いる方法、またはシーケンサーを用いて DNA の塩基配列を解読する方法により行うことができる。

【 0 0 5 7 】

(5) 透かし配列の耐性

第 1 の実施の形態と同様に、DNA 全体の複製、細胞移植による複製、染色体を抽出して行う複製、DNA 中の透かし配列を含む領域を特に切り出して行う複製などに対して、本実施の形態における透かし情報は耐性を持つ。

また、DNA から切り出した部分を複製した場合であっても、この部分に遺伝子が含まれているならば、そのイントロン部分も必ず複製される。したがって、遺伝子単位で複製が行われる限り、本実施の形態における透かし情報は耐性を持つ。遺伝子から蛋白質が合成される過程で蛋白質コード領域が転写されて mRNA となっている状態から複製された場合であっても同様である。

しかし、高等生物においては、イントロン部分は、1 次 mRNA からアミノ酸に翻訳される際にスプライシングにより切り落とされてしまうため、スプライシング後の mRNA からの複製に対しては、本実施の形態における透かし情報は耐性を持たない。

【 0 0 5 8 】

〔第 3 の実施の形態〕

次に、コドンの冗長性を利用して透かし情報を埋め込む方法における実施の形態を説明する。

上述したように、DNA の塩基配列は、3 文字からなるコドンを単位としてアミノ酸をコードする。しかし、生体内に存在する 20 種類のアミノ酸に対し、塩基 3 個では $64 (= 4^3)$ 通りの表現ができる。したがって、一つのアミノ酸に対するコドンのコードが複数存在する場合があります、この冗長部分に透かし情報を

埋め込む余地がある。

蛋白質合成の過程でアミノ酸に翻訳されるコドンとアミノ酸の対応関係はコドン表で与えられる。図 1 3 に示したコドン表を参照すると、一つのアミノ酸に対して複数のコドンが対応しており、その冗長性は主としてコドンの 3 文字目にあることが分かる。そこで、コドン表で許される範囲内、すなわち同一のアミノ酸に対応する限り、保護しようとする遺伝子のエキソン部分における各コドンを別の塩基に取り換える自由度があることになる。この自由度を利用して透かし情報を埋め込む。したがって、本実施の形態では、透かし情報を持たせるために選択されたコドンの配列が透かし配列を構成することとなる。

【 0 0 5 9 】

本実施の形態によれば、DNA の遺伝子部分のうち、エキソン部分に透かし情報を埋め込むことができるという利点がある。

図 1 1 は、本実施の形態による透かし配列の挿入方法の概念を説明する図である。

以下、本実施の形態に関して、図 2 に示した (1) 透かし配列の決定、(2) DNA への透かし配列の埋め込み、(3) 安全性の確認、(4) 透かし配列の検出の各ステップと、(5) 透かし配列の耐性について説明する。

【 0 0 6 0 】

(1) 透かし配列の決定

上述したように、一つのアミノ酸に対応する塩基配列のコード (コドン) が複数存在する。そのため、所定のアミノ酸に対応するコドンを意図的に選択することにより、有用な蛋白質をコードしている配列の意味 (コードされているアミノ酸配列) を変えることなく、付加情報を遺伝子に直接埋め込む事ができる。ここで、DNA 中の所望のコドン (またはコドンの一部の塩基) だけを書き換えるような厳密な書き換えは、現在では技術的に実現が難しいと思われる。

しかしながら、DNA の塩基配列を直接書き換えるのではなく、新しい蛋白質をアミノ酸レベルで設計し、または挿入したい外来の遺伝子のコードするアミノ酸配列を解読し、これに対応する DNA を設計する場合に、アミノ酸をコドンに読み換える段階で透かし情報を埋め込むことができる。

【 0 0 6 1 】

ただし、コドンの使い方は生物の種毎に差があり、通常偏っている。そのため、所定の生物において使用頻度の低いコドンを用いることは、対応する tRNA が少ないため、転写効率を低下させ、期待する蛋白質の機能が弱くなってしまうことが考えられる。

そこで、使用頻度が上位 2 つのコドンを使用し、これを 2 値データ (0、1) に対応させて情報を記述する方法を考える。

N 個のコドンを用いて各コドン (ただし、メチオニンに対応するコドンは 1 種類しかないのを除く) が 0 に対応するものか 1 に対応するものかを判別し、情報を読み取ると、長さ N の 2 値データのデータ列 (以下、ビット列) ができる。このビット列により透かし情報を記述する。

【 0 0 6 2 】

この方法を、さらに詳細に説明する。

蛋白質の合成効率になるべく影響を与えないようにするため、メチオニン以外の各アミノ酸に関して、それぞれ使用頻度の高い上位二つのコドンを選択し、“0”、“1”に割り当てる。情報を埋め込む遺伝子のエキソンに関して、コーディングに使用できる部分は全て使うことができる。また、疑似ランダム・キーに従って使用するコドン、使用しないコドンを区別し、検出時に同じキー (鍵) を用いて、埋め込みに用いられているコドンのみから情報を取り出すようにしても良い。

【 0 0 6 3 】

この方法を用いた場合に問題となる点が二つある。一つは、情報を埋め込んでいない個体の遺伝子からも、上述の規則に従って何らかのメッセージを取り出してしまうといった検出誤り (false positive error) が生じるおそれがある点である。すなわち、“1 0 0 1”というビット列が取り出された場合に、このビット列が意図的に埋め込まれた情報であるのか、単に偶然に生じた組み合わせなのかを区別する手段がない。

もう一つは、埋め込みを行った遺伝子において、何らかの変異が起こったために繰り返しの規則が壊れ、透かし情報を表すビット列を検出できなくなるような

検出誤り (false negative error) が生じるおそれがある点である。

【0064】

false positive errorによる問題点を解決する方法の一つとして、一つのメッセージを繰り返して埋め込む方法が考えられる。埋め込みを行っていない遺伝子において、そのような繰り返しが起こる確率が十分に低いならば、このビット列の繰り返しが検出されたDNAは透かし情報の埋め込みが行われたと判断することができる。

false positive errorが生じる確率は、次のようにして求められる。

まず簡単のため、コーディングに用いるアミノ酸をAの1種類とする。このアミノ酸Aを合成するコドンのうち、個体中での使用頻度が最大のものをC0（使用確率P0）、使用頻度が次に大きいものをC1（使用確率P1）とする。ビットの割り当ての例としては、C0を“0”、C1を“1”とする。また、埋め込みを行う遺伝子のエキソン中に含まれるC0、C1の数を合計でNとする。ここで、上述の通りC0、C1の組み合わせによる所定のビット列で表される透かし情報をm回繰り返して埋め込む。この場合、nビット（ $N = mn$ とする）の情報を埋め込むことが可能である。

以上の仮定の下でfalse positive errorが生じる確率は次式（数1）で表される。

【0065】

【数1】

$$(\text{false_positive_error}) = \sum_{k=0}^n \binom{n}{k} (P0)^k (P1)^{n-k}$$

$$\text{ただし、} \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

【0066】

コーディングに用いるアミノ酸を複数とした場合であっても、エキソン中での各アミノ酸の出現頻度を上式（数1）に適用して計算することができる。

【0067】

また、上述した n ビットのうち、 s ビット ($s < n$) をメッセージに用い、残りの $(n - s)$ ビットをエラー訂正符号として用いることによって、false negative error の確率も十分に減らすことができる。

【0068】

なお、以上の説明では、アミノ酸 A を合成するコドンのうち出現頻度が上位二つのコドン C0、C1 に対して 0、1 のビットを割り当てることとしたが、同じアミノ酸 A を合成する C0、C1 以外のコドンも全て 0、1 に置き換えることによって、埋め込む情報量を増やすことができる。

【0069】

(2) DNA への透かし配列の埋め込み

遺伝子を合成する過程で、特定のコドンを構成する塩基を適宜選択し、透かし情報を表すビット列を作成することにより、DNA に透かし情報を埋め込むことができる。また、遺伝子の合成技術の延長として、1 塩基単位で置き換え操作を行ったり、1 コドン単位で塩基の置き換え操作を行うことにより、同様に DNA に透かし情報を埋め込むことができる。

【0070】

(3) 安全性の確認

本実施の形態により透かし情報を埋め込まれた遺伝子は、対象である個体が本来持っている遺伝子と同一の蛋白質を合成する遺伝子ではあるが、遺伝子中の各コドンが冗長性の範囲内で人工的に書き換えられているので、個体に対して副作用が全くないとは言い切れない。したがって、本実施の形態においても、透かし情報が埋め込まれた個体の安全性を確認することが必要である。

安全性の基準は、上述した第 1 の実施の形態と同様に、保護しようとする付加価値のある遺伝子の機能に応じて決められるべきである。

また、安全性の確認には、生物を用いた実験による安全性のテストが行われるべきである。透かし配列の安全性を確認する手順は、第 1 の実施の形態における図 7 に示した手順と同様である。

【0071】

(4) 透かし配列の検出

透かし配列の検出は、第 1 の実施の形態の場合と同様に、埋め込んだ透かし配列に相補的な塩基配列を用いる方法、またはシーケンサーを用いて DNA の塩基配列を解読する方法により行うことができる。

【 0 0 7 2 】

(5) 透かし配列の耐性

第 1、第 2 の実施の形態と同様に、DNA 全体の複製、細胞移植による複製、染色体を抽出して行う複製、DNA 中の透かし配列を含む領域を特に切り出して行う複製などに対して、本実施の形態における透かし情報は耐性を持つ。

また、第 2 の実施の形態と同様に、DNA から切り出した部分を複製した場合であっても、この部分に遺伝子が含まれているならば、そのイントロン部分も必ず複製される。したがって、遺伝子単位で複製が行われる限り、本実施の形態における透かし情報は耐性を持つ。遺伝子から蛋白質が合成される過程で蛋白質コード領域が転写されて mRNA となっている状態から複製された場合であっても同様である。

さらに、本実施の形態は、遺伝子のエキソン部分に透かし情報を埋め込むため、最終的にアミノ酸に翻訳される段階の mRNA にも透かし情報が含まれることとなる。そのため、スプライシング後の mRNA からの複製に対しても、本実施の形態における透かし情報は耐性を持つ。

【 0 0 7 3 】

以上のようにして、本発明の各実施の形態により DNA に埋め込まれた透かし情報は、それぞれの耐性に応じて遺伝情報の出所を判定するための情報として用いることができる。

図 1 2 は、複製方法ごとに各実施の形態の耐性を示した図表である。

図 1 2 を参照すると、交配に対しては、第 1、第 2、第 3 の実施の形態による透かし配列の全てが耐性を有する。1 次 RNA の複製に対しては、第 2、第 3 の実施の形態による透かし配列が耐性を有する。スプライシング後の複製に対しては、第 3 の実施の形態による透かし配列が耐性を有する。

【 0 0 7 4 】

この透かし配列を検出し解析することにより、透かし配列と共に当該遺伝子中

に含まれる付加価値のある遺伝子が特定の遺伝子の複製であることを確認することができる。さらに、この付加価値のある遺伝子を生産したり複製したりする権利を契約その他の手段で制限している場合は、その遺伝子の複製が正当なものかどうかを判断し、不正な複製を防止することが可能となる。

【 0 0 7 5 】

【発明の効果】

以上説明したように、本発明によれば、DNAの塩基配列に所定の情報を埋め込むことにより、当該情報を持つDNAにおける遺伝情報の出所を判定することが可能となる。

【 0 0 7 6 】

また、本発明によれば、DNAの塩基配列に意図的に埋め込まれた情報を検出し解析することにより、所定の個体が持つ所定の遺伝子が特定の遺伝子の複製かどうかを判断することが可能となる。

【 0 0 7 7 】

さらにまた、本発明によれば、所定の個体が持つ所定の遺伝子が特定の遺伝子の複製かどうかを判断可能とすることにより、第三者による特定の遺伝子の不正な複製を防止する手段を提供することができる。

【図面の簡単な説明】

【図 1】 DNA中の遺伝子から蛋白質が合成される過程を説明する図である。

【図 2】 本発明により透かし配列を決定してDNAに埋め込み、検出するまでの全体的な処理の流れを説明するフローチャートである。

【図 3】 本発明の第 1 の実施の形態による透かし配列の挿入方法の概念を説明する図である。

【図 4】 本実施の形態において、DNAシーケンスデータを基にしてこの確率の計算を行い、透かし配列の候補を決定する手順を示すフローチャートである。

【図 5】 疑似データを用いて 1 個体中における長さ 6 塩基の塩基配列の度数分布の例を示す図である。

【図 6】 図 5 における塩基配列 A A A G T C の出現頻度に対する個体数の度数分布表の例を示す図である。

【図 7】 本実施の形態における透かし配列の安全性を確認する手順を示すフローチャートである。

【図 8】 相補的な塩基配列を用いて透かし配列を検出する様子を示す図である。

【図 9】 シークエンサーを用いて D N A の塩基配列を解読し、透かし配列を検出する様子を示す図である。

【図 1 0】 本発明の第 2 の実施の形態による透かし配列の挿入方法の概念を説明する図である。

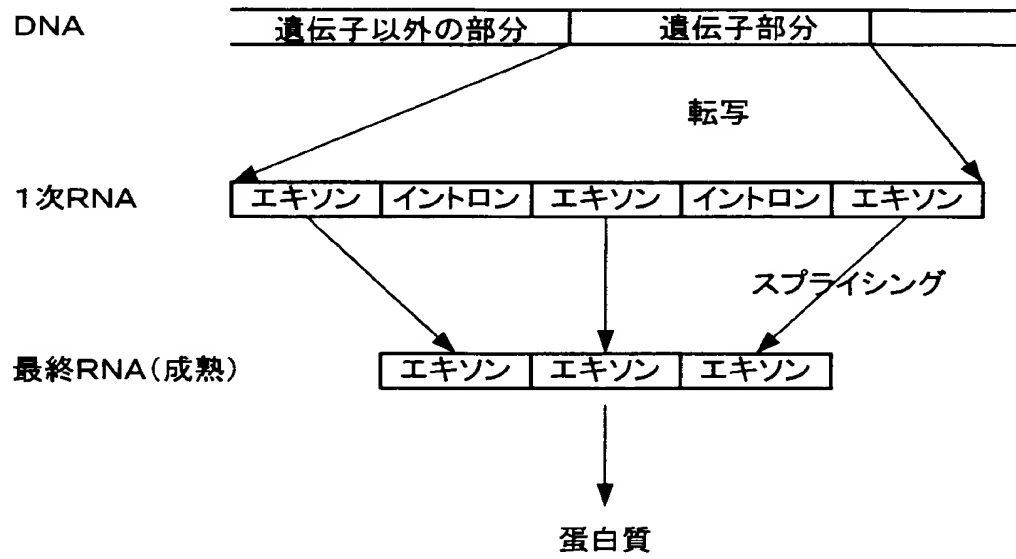
【図 1 1】 本発明の第 3 の実施の形態による透かし配列の挿入方法の概念を説明する図である。

【図 1 2】 複製方法ごとに各実施の形態の耐性を示した図表である。

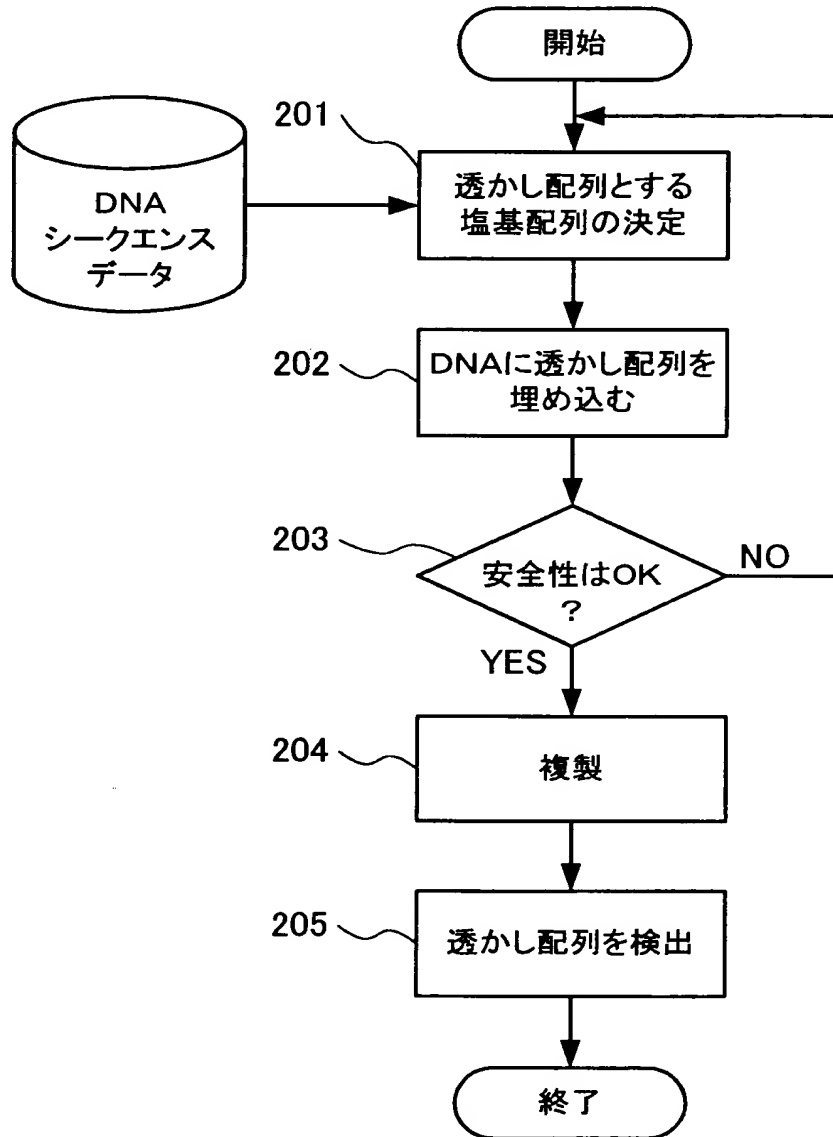
【図 1 3】 コドンとそれに対応するアミノ酸（または特別な意味）をまとめた図表である。

【書類名】 図面

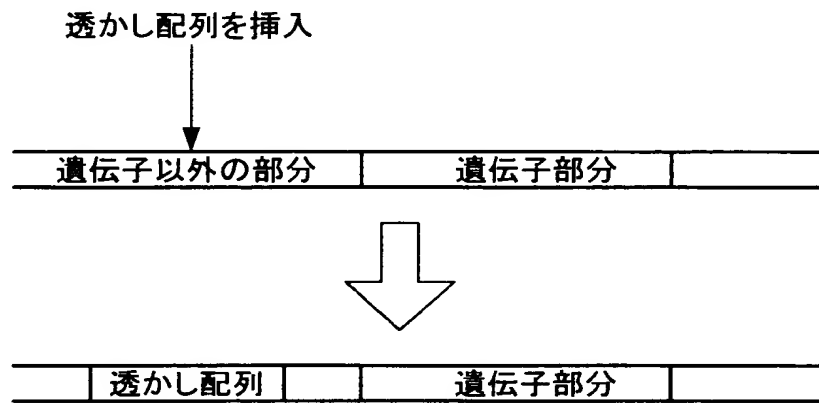
【図 1】



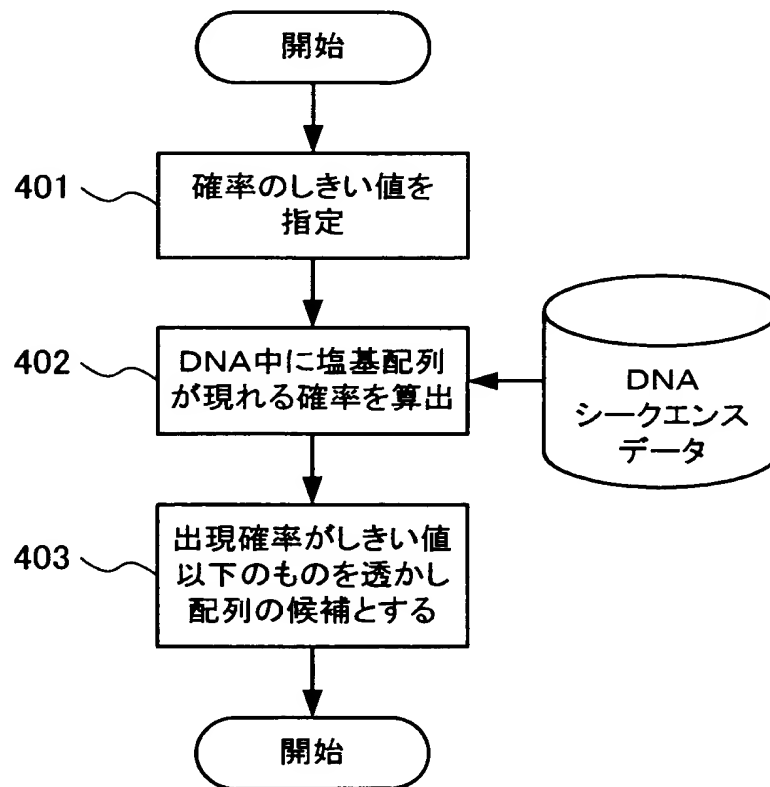
【図 2】



【図 3】



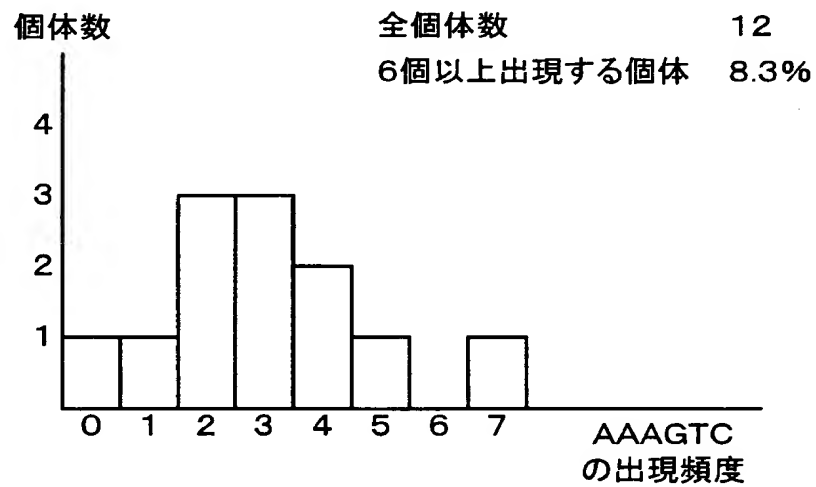
【図 4】



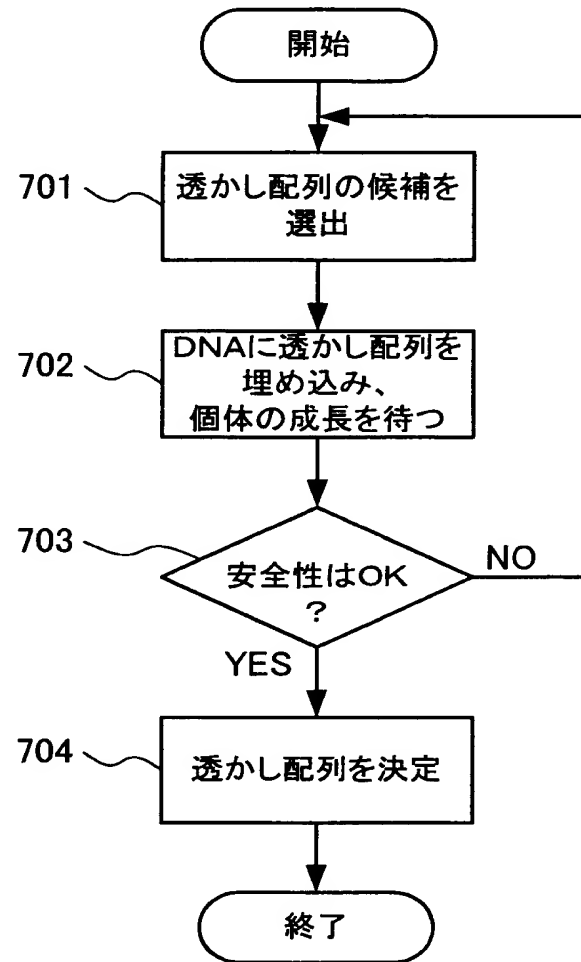
【図 5】

長さ6塩基の部分配列	ある個体での頻度
AAAGTT	12
AAAGTG	50
AAAGTC	3
AAAGTA	11
AAAGGT	2
AAAGGG	0
...	...

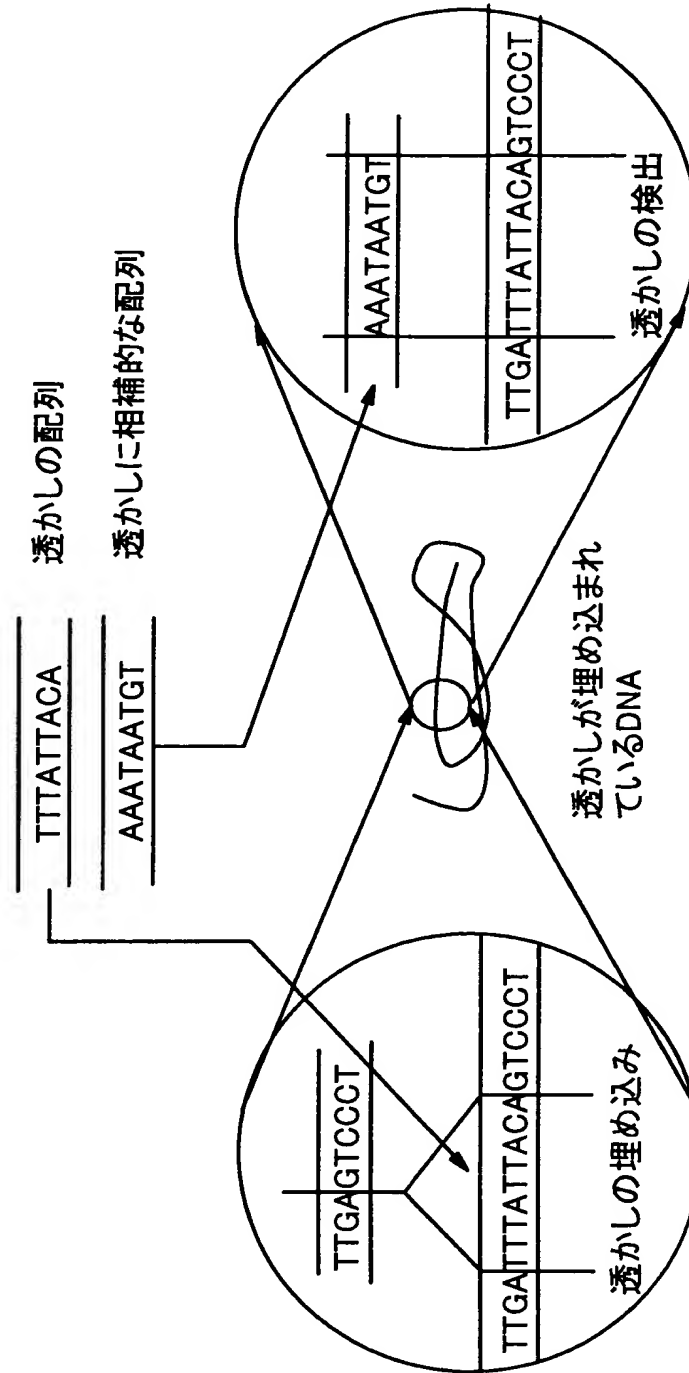
【図 6】



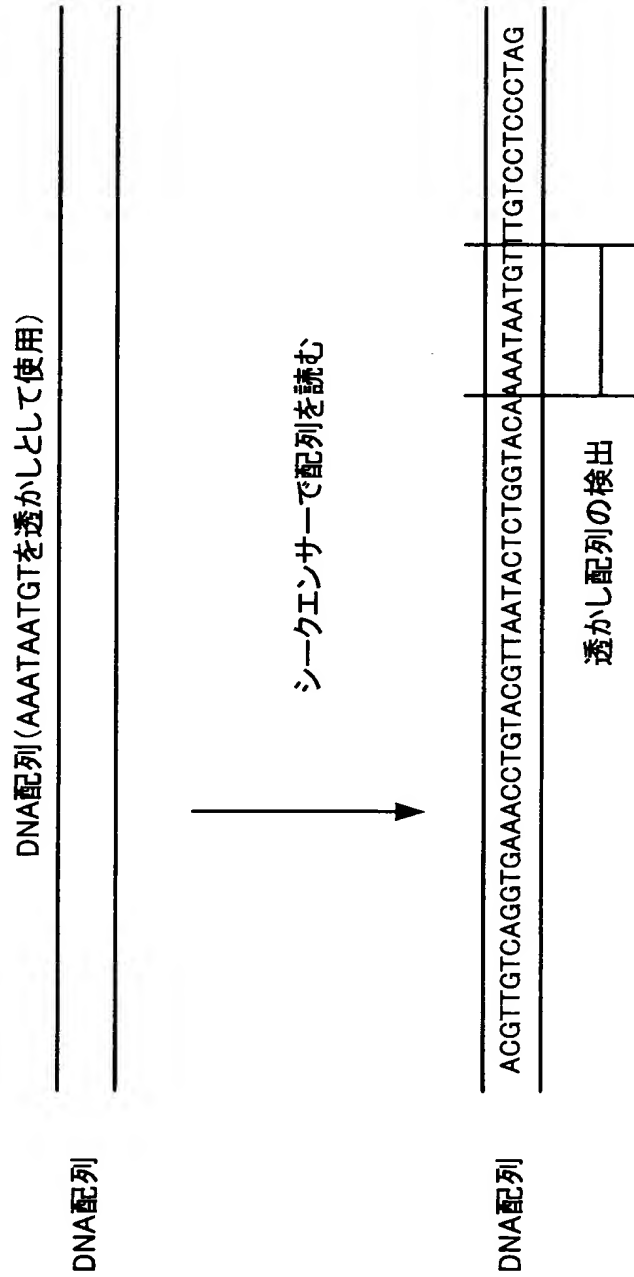
【図 7】



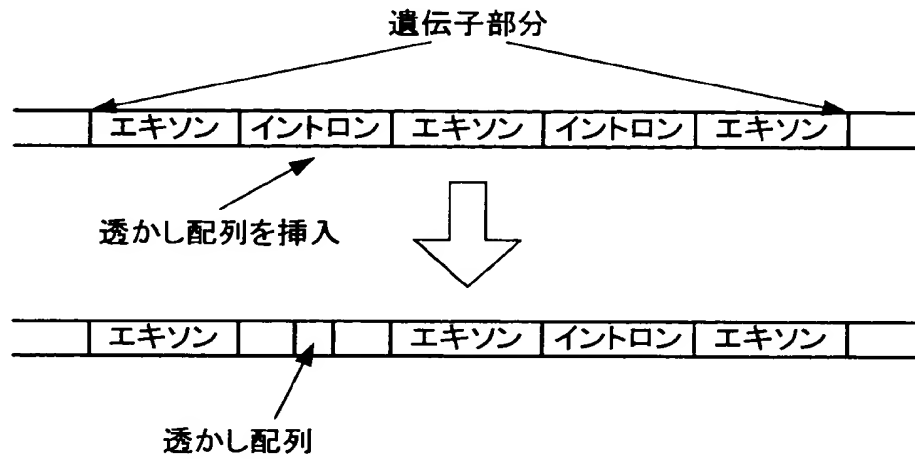
【図 8】



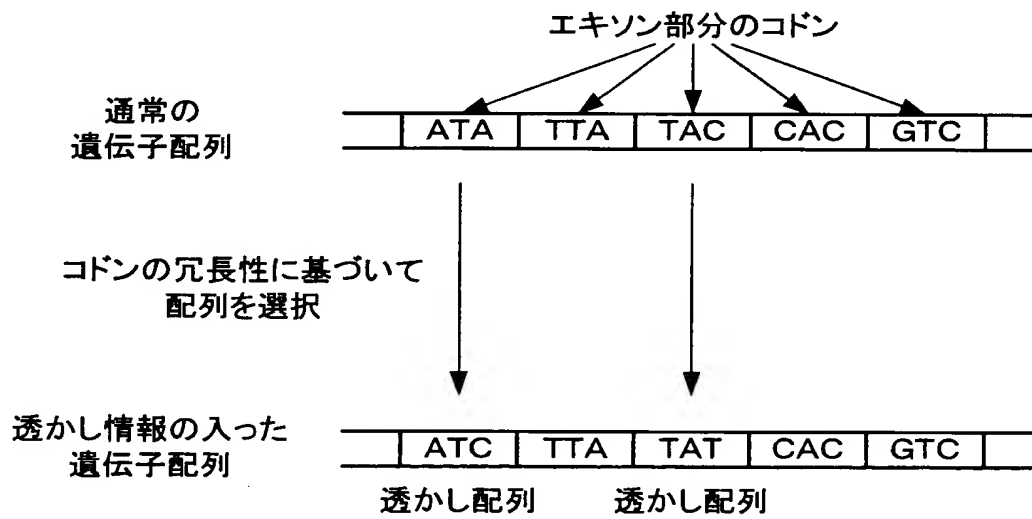
【図 9】



【図 10】



【図 11】



【図 1 2】

	第1の実施の形態 (遺伝子以外の部分 に透かし配列を挿入)	第2の実施の形態 (イントロンに透かし 配列を挿入)	第3の実施の形態 (コドンの冗長性を利用 して透かし情報を挿入)
交配に対する耐性	有り	有り	有り
1次RNAの複製 に対する耐性	無し	有り	有り
スプライシング後の 複製に対する耐性	無し	無し	有り

【図 1 3】

UUU Phe UUC Phe UUA Leu UUG Leu	UCU Ser UCC Ser UCA Ser UCG Ser	UAU Tyr UAC Tyr UAA 終止 UAG 終止	UGU Cys UGC Cys UGA 終止 UGG Trp
CUU Leu CUC Leu CUA Leu CUG Leu	CCU Pro CCC Pro CCA Pro CCG Pro	CAU His CAC His CAA Gln CAG Gln	CGU Arg CGC Arg CGA Arg CGG Arg
AUU Ile AUC Ile AUA Ile AUG Met	ACU Thr ACC Thr ACA Thr ACG Thr	AAU Asn AAC Asn AAA Lys AAG Lys	AGU Ser AGC Ser AGA Arg AGG Arg
GUU Val GUC Val GUA Val GUG Val	GCU Ala GCC Ala GCA Ala GCG Ala	GAU Asp GAC Asp GAA Glu GAG Glu	GGU Gly GGC Gly GGA Gly GGG Gly

【書類名】 要約書

【要約】

【課題】 D N A の塩基配列に所定の情報を埋め込むことにより、当該情報を持つ D N A における遺伝情報の出所を判定できるようにする。

【解決手段】 D N A 中に通常現れないパターンの塩基配列と、この D N A が有する所定の遺伝情報の出所を識別するための識別情報とを対応付け、この D N A に対して、この D N A の遺伝情報に影響を与えないように、この識別情報に対応付けられた塩基配列を埋め込む。

【選択図】 図 3

認定・付加情報

特許出願の番号	特願 2 0 0 0 - 1 7 1 0 3 0
受付番号	5 0 0 0 0 7 0 8 1 2 8
書類名	特許願
担当官	東海 明美 7 0 6 9
作成日	平成 1 2 年 7 月 1 9 日

<認定情報・付加情報>

【特許出願人】

【識別番号】	390009531
【住所又は居所】	アメリカ合衆国 1 0 5 0 4、ニューヨーク州 アーモンク (番地なし)
【氏名又は名称】	インターナショナル・ビジネス・マシーンズ・コーポレーション

【代理人】

【識別番号】	100086243
【住所又は居所】	神奈川県大和市下鶴間 1 6 2 3 番地 1 4 日本アイ・ビー・エム株式会社 大和事業所内
【氏名又は名称】	坂口 博

【復代理人】

【識別番号】	申請人
【識別番号】	100104880
【住所又は居所】	東京都港区赤坂 7 - 1 0 - 9 第 4 文成ビル 2 0 2 セリオ国際特許事務所
【氏名又は名称】	古部 次郎

【選任した代理人】

【識別番号】	100091568
【住所又は居所】	神奈川県大和市下鶴間 1 6 2 3 番地 1 4 日本アイ・ビー・エム株式会社 大和事業所内
【氏名又は名称】	市位 嘉宏

【選任した復代理人】

【識別番号】	100100077
【住所又は居所】	東京都港区赤坂 7 - 1 0 - 9 第 4 文成ビル 2 0 2 セリオ国際特許事務所
【氏名又は名称】	大場 充

次頁無

出 願 人 履 歴 情 報

識別番号 [390009531]

1. 変更年月日 2000年 5月16日
[変更理由] 名称変更
住 所 アメリカ合衆国10504、ニューヨーク州 アーモンク (番地なし)
氏 名 インターナショナル・ビジネス・マシーンズ・コーポレーション